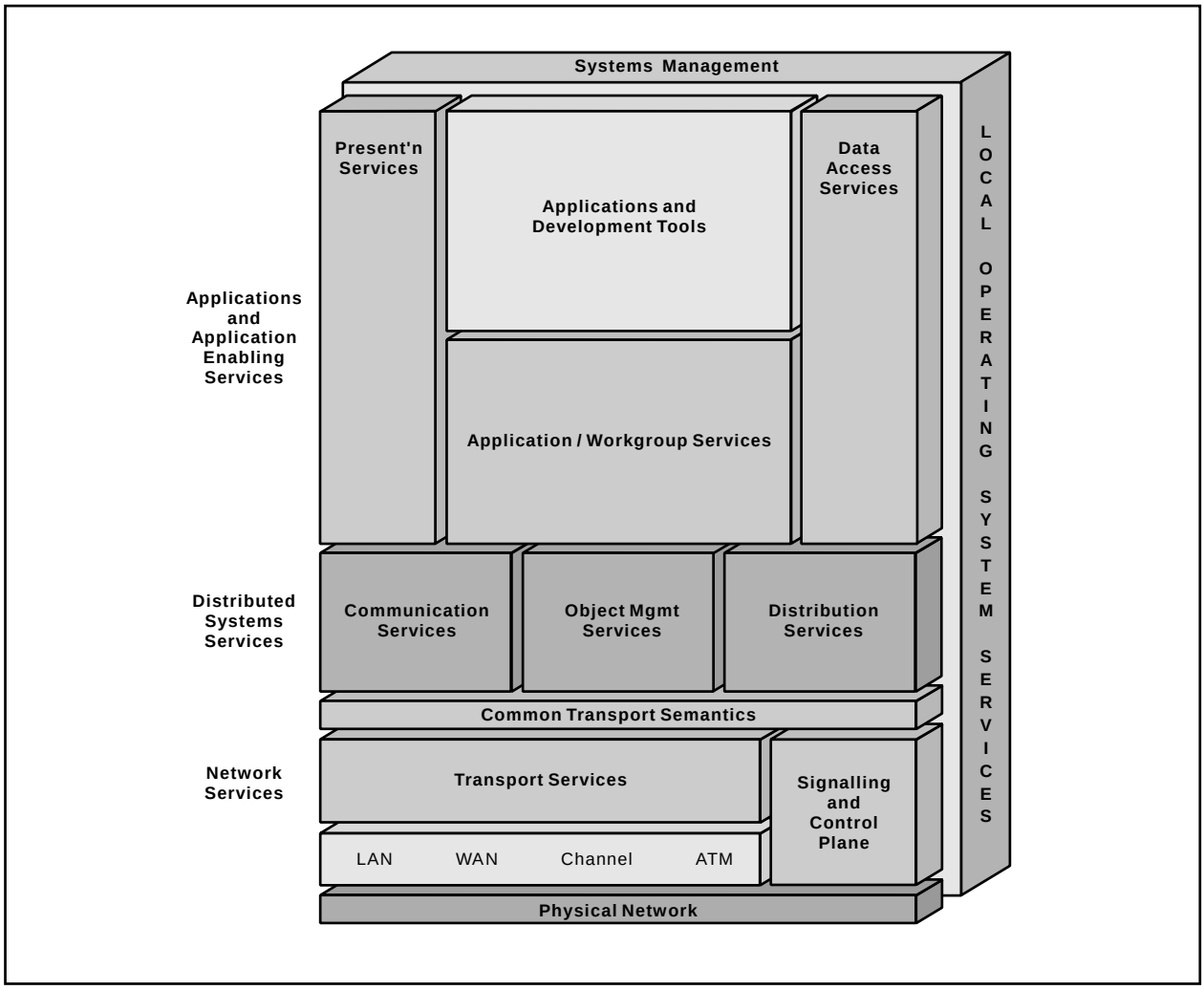
Open Blueprint

Digital Library Resource Manager



GC23-3909-01

Open Blueprint

Digital Library Resource Manager

GC23-3909-01

About This Paper

Open, distributed computing of all forms, including client/server and network computing, is the model that is driving the rapid evolution of information technology today. The Open Blueprint structure is IBM's industry-leading architectural framework for distributed computing in a multivendor, heterogeneous environment. This paper describes the Digital Library Resource Manager component of the Open Blueprint and its relationships with other Open Blueprint components.

The Open Blueprint structure continues to accommodate advances in technology and incorporate emerging standards and protocols as information technology needs and capabilities evolve. For example, the structure now incorporates digital library, object-oriented and mobile technologies, and support for internet-enabled applications. Thus, this document is a snapshot at a particular point in time. The Open Blueprint structure will continue to evolve as new technologies emerge.

This paper is one in a series of papers available in the *Open Blueprint Technical Reference Library* collection, SBOF-8702 (hardcopy) or SK2T-2478 (CD-ROM). The intent of this technical library is to provide detailed information about each Open Blueprint component. The authors of these papers are the developers and designers directly responsible for the components, so you might observe differences in style, scope, and format between this paper and others.

Readers who are less familiar with a particular component can refer to the referenced materials to gain basic background knowledge not included in the papers. For a general technical overview of the Open Blueprint, see the *Open Blueprint Technical Overview*, GC23-3808.

Who Should Read This Paper

This paper is intended for audiences requiring technical detail about the Digital Library Resource Manager in the Open Blueprint. These include:

- Customers who are planning technology or architecture investments
- Software vendors who are developing products to interoperate with other products that support the Open Blueprint
- Consultants and service providers who offer integration services to customers

Contents

Figures

Open Blueprint Digital Library Resource Manager

What is a Digital Library?

Traditional libraries collect and store information, cataloged so that the information can be found and retrieved easily. These cataloged materials are called holdings. After a holding is acquired, cataloged, and added to a collection, it is made available to authorized users. Measures are put in place to preserve and protect the holdings.

An application of a digital library follows essentially the same model as illustrated in Figure 1. The digital library server manages the catalog as well as the collection of holdings. The digitized collection of the holdings is a digital analog of the library stacks.

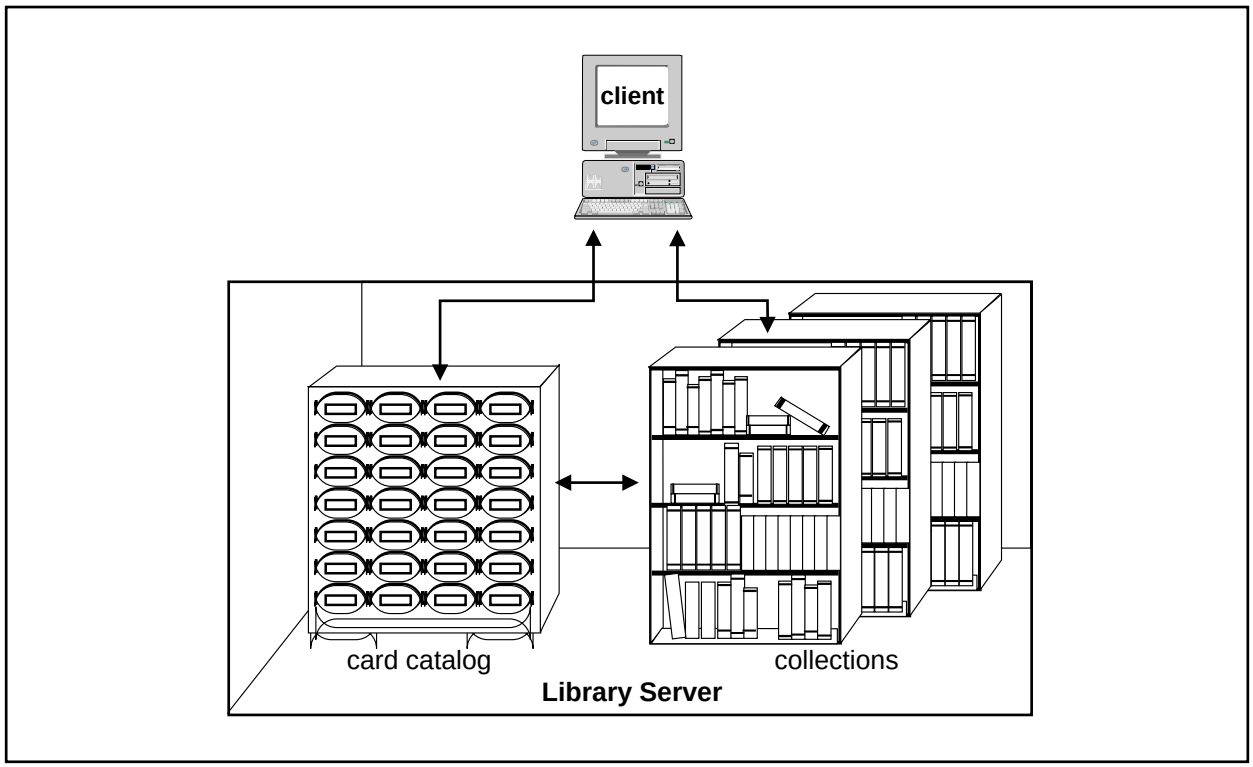


Figure 1. A Simple Digital Library

Digital libraries contain widely disparate types of information: text, images, video, and so on. The holdings in a library have different values, origins, and longevities. The users of a digital library have different purposes as well as different operating environments. Digital libraries serving 100,000 students on a multicampus university have different requirements than libraries that serve 200 film editors in a movie studio, or libraries that provide business information to commercial customers.

A complete digital library solution integrates extensive information capture, storage and management of information, intellectual property rights management, search and access, and distribution technologies into a single library architecture. Figure 2 on page 2 illustrates the scope of a complete digital library solution.

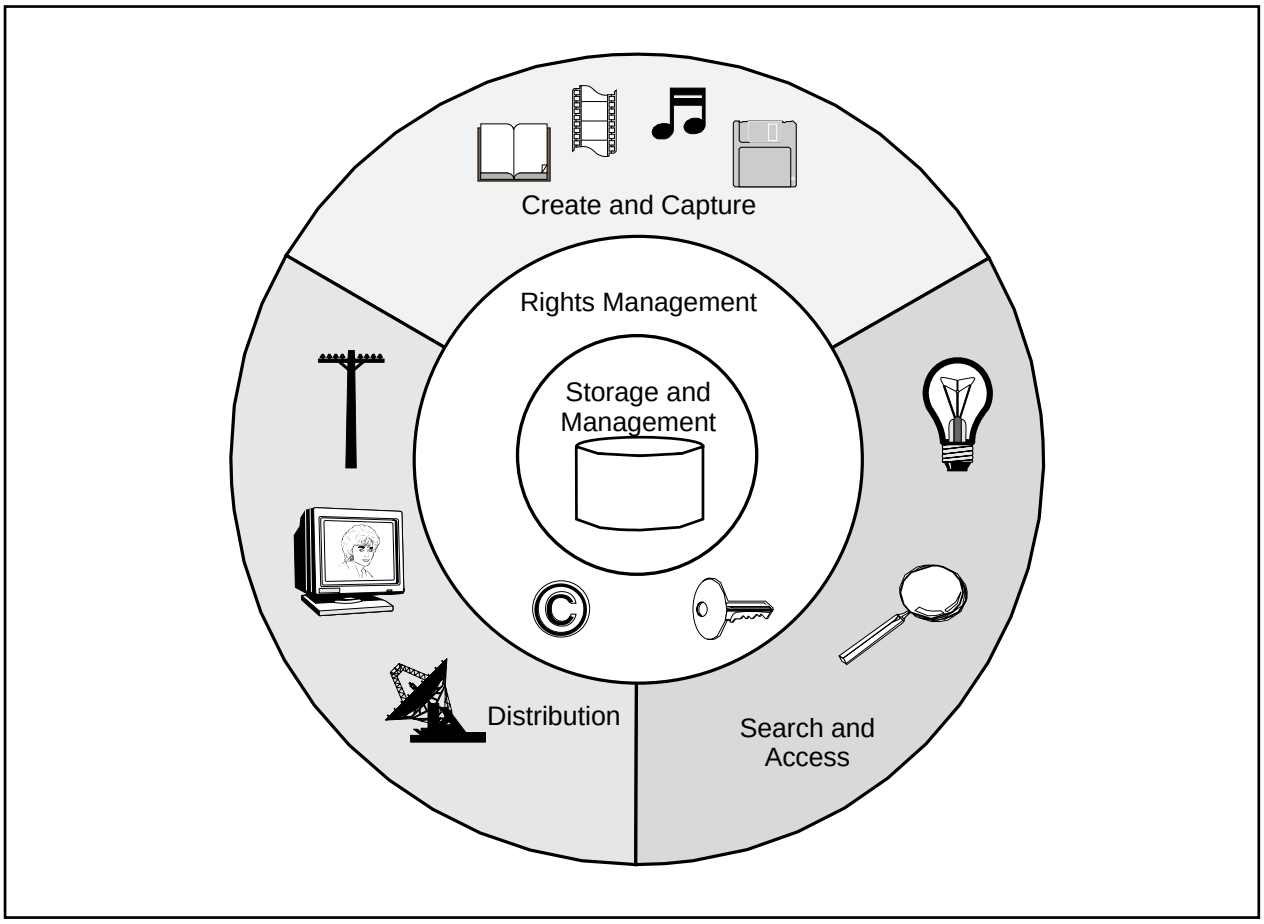


Figure 2. Scope of a Digital Library Solution

What is the Digital Library Resource Manager?

The Digital Library resource manager provides the programming interfaces, class libraries, and library administration facilities to build digital libraries that scale from solutions for small in-house libraries, to multicampus libraries, to globally interconnected libraries. The Digital Library resource manager is the enabling software that allows the integration of the required technologies into a digital library solution that addresses the specific requirements of the library and the users of the library. The Digital Library resource manager itself does not contain any library holdings.

The Digital Library resource manager is a distributed resource manager in the Open Blueprint. The Digital Library resource manager mimics the traditional library functions of preserving and protecting holdings and enabling the payment of royalties where applicable, as well as providing digital storage and access functions.

Terminology

The following terms are defined for this paper:

Collection A *collection* is an aggregate of holdings, gathered together because of one or more shared attributes. Also called a folder in some digital library implementations. A collection may contain other collections as well.

Content *Content* refers to the digitized information, such as the bitmap of a picture, the digitized text of a document, the digitized video, and so on.

Holding A *holding* is an individual intellectual property that may consist of one or more parts. The holding includes its parts, rules about access and usage, and searchable attributes that apply to the holding and its parts. Attributes include information kept in the catalog, as well as the actual digitized content (if any) of the holding. A holding is the smallest cataloged unit in the library.

A holding is referred to as an *item* by the Digital Library resource manager's interface definition.

Library A *library* refers to an instance of a digital library solution, including the client access, the library servers and the holdings.

Library client The *library client* is the software that gives users access to the holdings in a library.

Library server A *library server* consists of a catalog, collections of holdings, and server support for all the digital library technologies as illustrated in Figure 2 on page 2. These components can be distributed among one or more systems.

Part A *part* is a unit of intellectual property contained within a holding, and independently presented. For example, a periodical is a holding that contains text and pictures, any one of which may be a separate intellectual property with its own license agreement covering access, use and compensation.

A part is accessed only through the holding of which it is a part.

How the Digital Library Resource Manager Supports Digital Library Solutions

The Digital Library resource manager enables the development of complete digital library solutions by supporting these functions:

- Content creation and capture
- Storage and management
- Distribution
- Search and access
- Rights management

Content Creation and Capture

Digital information is created on a computer from a wide variety of authoring, painting, drawing and editing tools, or captured from non-digital sources using a variety of capture tools and applications. These tools and applications can be integrated into a digital library solution through the Digital Library resource manager's open programming interface. These content capture and creation tools can access digital libraries for existing raw material or content to work with, or they can place finished products directly into the library.

After the physical media have been scanned or transformed into digital format, or created digitally, they must be cataloged and indexed so that they can be effectively searched. Significant attributes about the content are extracted from the digitized holdings and enriched with other user-provided attributes that can be used as finding aids. For example, features like color, shape, texture, and their position within the image are extracted from images and stored as attributes to be examined in subsequent searches.

Many types of library holdings do not warrant creating a rigorous catalog entry, since it takes time for an experienced cataloger to generate a single entry from scratch. Libraries with a few users and a large number of holdings could find this particularly burdensome.

Therefore, the Digital Library resource manager automates the work of cataloging the holding by extracting some catalog data from the digitized content. It supports whatever level of rigorous cataloging and indexing is required by the library.

Storage and Management

The holdings in a digital library are valuable assets of the creators, and of the library owners who make the holdings available to others. The Digital Library resource manager must provide reliable storage and management of the holdings to ensure their integrity, both in the creators' libraries, and in the libraries of any licensees. The Digital Library resource manager uses open and scalable storage and management systems that can support the volume, transaction rates, and storage technology required by the library solution.

Library client software can be anywhere users wish it to be, and are not depended upon to protect data held by the library. Instead, the library server is responsible for the physical protection of the information in the library. The physical protection is provided by the storage and management functions of the Digital Library resource manager. This means that the library must be installed by trusted administrators in secure locations.

In a library, the smallest cataloged unit is the holding. A holding has a unique identifier that is used to retrieve the materials from the library.

Any type of holding can be managed by the storage and management subsystem. The library records whatever information is needed about the holding and its relationships.

Holdings can be more than one physical information object, or part. For example, a set of slides is a holding for cataloging and distribution purposes, but each slide is a separate part of the holding that can be separately presented. A book is another example of a holding. Books may contain parts such as text, images, or hyperlinks, each part governed by a different set of presentation, permissions, and compensation rules.

A part is usually a single unit. Some parts, however, can be stored as many different renderings, to support efficient display on a variety of monitors, or different uses of the part (for example, a visitor casually viewing a fossil image at a tourist kiosk, versus the serious paleontologist viewing the same image). A part is the smallest independently stored and presented information object.

For the Digital Library resource manager, the holding is the unit used to link together the digitized parts of one intellectual property, as well as all the information needed to present them, and the rules for managing their rights and permissions.

The Digital Library resource manager optimizes the use of library resources by moving seldom used holdings to more cost-effective secondary storage. It automatically recalls these migrated holdings when they are accessed.

Distribution

Figure 3 on page 5 shows the potential distribution path for library holdings. Experience with different library solutions identifies four kinds of participants: authors or creators (source libraries), editors and publishers, subscribers (custodial libraries), and users. The information flows and the responsibilities of each library server in the path emulate traditional libraries. Each library server represents an administrative domain, not necessarily a separate computing system.

The relationship between the custodial, or campus, libraries and the authors is many-to-many. Authors can license their material to many publishers, and publishers can license material from many authors.

Users normally access custodial libraries, but in some cases, they might actually access the source library itself. Each of the pathways represented imply potentially different controls and payment flows. The Digital Library resource manager can accommodate digital library solutions for many application scenarios.

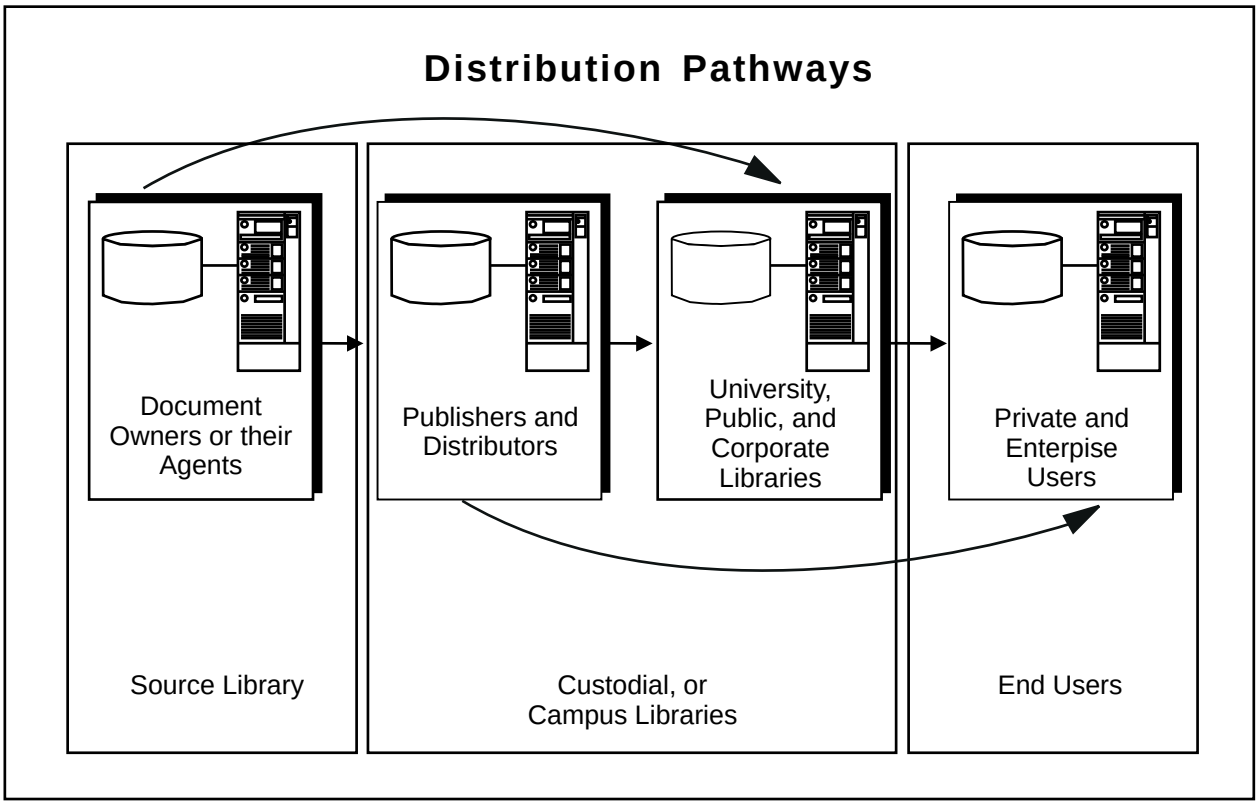


Figure 3. Library Distribution Pathways

Source libraries and custodial libraries can be supported across a variety of network configurations, from stand-alone systems to local area networks (LANs), wide area networks (WANs), or the Internet. Three types of delivery environments are considered:

1. World Wide Web (WWW) browsers (Hypertext Transfer Protocol (HTTP) clients)
2. Collaborative work-group clients (applications using the Collaboration resource manager)
3. Custom clients, using the Digital Library resource manager library client running in the users' workstations.

Public and private enterprises can implement a digital library and make its content available to users on existing computer networks or commercial online services like IBM Global Network or Prodigy, or public

networks like the Internet. See “Distributed Libraries” on page 9. Universities, public libraries, music distribution companies, cultural institutions, and other enterprises are also implementing digital libraries.

Library information can be presented or delivered in many different forms. For example, users might find what they are looking for by browsing abstracts or thumbnail images. After they find what they want, they can order the complete version. The library might deliver the materials on a CD-ROM, floppy disks, on paper, or it might deliver the materials through a network.

Digital information can be very large in size, putting a strain on many existing network capabilities. Not only is sufficient bandwidth required, but real-time delivery of audio and video are dependent on a paced, continuous delivery of data over the network.

The holdings in the library are made available by store and forward techniques, or streamed to the client using the Multimedia resource manager. For store and forward delivery, holdings are sent directly from the collections to the client application. For stream delivery, such as for on-demand video or audio, a client player method is invoked by the client application using the object handle returned in the search results. The handle will be provided to the Multimedia resource manager for streaming from the collection to the destination devices.

In some cases, an intervening purchase decision is necessary before a library holding is made available to the client.

Search and Access

The best search methods depend upon the type and size of the data and the user's objectives. Thus, a portfolio of search technologies is used by the Digital Library resource manager to provide the widest possible choice of search methods.

Text search includes a free-text query facility that allows users to express queries in a natural style without worrying about exact word positioning. The result of a text search is a list of documents sorted by highest probability of relevance. Probabilistic formulae are used to compute the rankings.

Textual analysis of words is also performed. Proper nouns like the White House are distinguished from a common noun like white house. Concept-based text retrieval makes use of natural language processing techniques with statistical linguistic and semantic evidence to rank documents.

Clustering techniques organize the results of a search so that holdings with similar content are grouped together for easier selection. Abstracting reduces the search results to its main points, so that the user does not have to view all of the text to locate desired information.

Query by image content is another search technique used by the Digital Library resource manager. Query by image content allows a collection of images to be searched on the basis of colors, shapes, textures, and their positions within the image.

Queries can be based on the properties of an outlined object in the image, or they can be based on the entire image. For example, the query can be based on the car in a picture, or it can be based on the entire picture. The query parameters can be specified visually, for example, by selecting a color from one image to search for other images with the same color. Search results can be refined with subsequent queries to narrow down the results. Results are returned with a relevance ranking based on closeness to the specified parameter.

Each type of search (text, query by image content, or other parametric data) uses different search server support. Combined queries will use multiple search servers.

Combined queries can result in long results lists which affect performance and scalability in a client. Therefore, the Digital Library resource manager query support is designed so that functions can be divided between the client and a server.

The server part of the query support handles the combination of large results lists and rights management issues (for example, excluding from a results list those holdings that a user is not allowed to access), and will provide a place for agents to run (for example, scheduled queries, comparing new library additions to a user profile, and so on).

The Digital Library resource manager can be adapted to support additional search technology as it becomes available. Future developments in video indexing techniques will detect properties such as scene changes and other video metrics. These techniques will allow searching video streams, as well as indexing to particular scenes in a video.

Rights Management

Network computing and digital technology make information widely available over a variety of networks, from LANs to the Internet. With data so widely available, owners of intellectual property expect some control over how their property is used. They are concerned about unauthorized use or redistribution of their materials when they release these materials to a digital library.

But, rights management is not only about protecting holdings and compensating owners. Equally important is protecting users from counterfeit or corrupted material. For example, the academic and publishing community is concerned about the mass distribution of data that has not been editorially scrutinized or subjected to peer review.

One way that owners of intellectual property can exercise control over how their property is used is through license agreements and compensation rules. Figure 3 on page 5 shows how intellectual property might flow from one library to another, each path involving potentially unique agreements. The rights and royalties agreements that apply when an intellectual property is transferred from the source library to one or more custodial libraries are different from the rights and royalties that apply when the property is distributed from the custodial libraries to the users. The relationships between source libraries, custodial libraries, and users are many-to-many, and each instance of a relationship may have unique intellectual property rights management requirements.

An example is the licensing of a film by a producer to a cable TV company. The licensing and billing policies that apply between the producer and the cable company are quite different from the licensing and billing policies that apply between the cable company and the subscriber.

Because holdings are made up of embedded parts, multiple licensing, rights protection, and usage fees may apply. The pictures in a book, for example, may be the property of the photographer or painter, while the book copyright is owned by the author.

Rights management means that national and international laws relating to intellectual property, such as copyrights, patents, and trademarks are followed. Rights management also applies to contractual obligations between content owners, custodial libraries, and users. Library holdings must be protected not only in the server, but also after they have been transmitted. Therefore, workstations must contain trusted programs that can securely receive the holding, and can prevent or report unauthorized uses. These trusted programs are management compliant and work with the Digital Library resource manager to protect intellectual property rights. Rights management compliant browsers are downloaded to user workstations when required and must execute immediately, or before their expiration time elapses.

A rights management launch pad will be part of the Digital Library resource manager to support WWW browsers.

Solutions to intellectual property rights management must account for different problem areas:

- Identification and authentication

  Identification and authentication of library users is important with respect to individual access to the information, especially when payment, whether by subscription, membership, or fee-based, is required. Equally important, library users want to be sure that the materials they receive are authentic.

  The Digital Library resource manager ties access control to parts of holdings, not just to a single holding. This means that access control, as well as metering and rights management, can be applied individually to each part of a holding.

  From time to time, the identity of a user or a key that identifies a group of users, might be needed. In some library applications, the content provider or library owner may require a separate logon and password to identify users. Other libraries might find this unnecessary or burdensome due to a rapid turnover in library users or because the intent of the library is to make information widely available to the public. In this case, libraries will elect to inherit the user authentication from the client system. The certificate that identifies and authenticates the user will be obtained from the Identification and Authentication resource manager.

- Protection

  Reasonable techniques to avoid misuse of assets and lost royalties must be employed by the library system.

  License agreements between the authors of holdings and custodial libraries, and between custodial libraries and subscribers, must be in place before the holding is made available to users. These agreements provide the legal language defining rules for storage, use, and access to the digitized information, as well as royalties or other compensation to the rights holder. Licensing agreements, rights, and royalties are often individually negotiated, and can be different depending on the distribution medium and the type of users that the library serves. Also, the rights to distribute or to use the intellectual property might be good only for a specific time period.

  Measures designed to protect library holdings from unauthorized use or redistribution can take into account the value of the holding. More stringent measures can be employed for high value holdings, or confidential holdings, than the measures employed for low value holdings. Different parts of a holding can have different levels of protection. For example, an abstract can be transmitted unencrypted to all users while the main body of text is protected.

  Two techniques which can be used are:

  – Marking

    Visible and invisible marking technology can be integrated into a digital library solution.

    One technique provided by the Digital Library resource manager is a visible watermark for marking digitized images. The watermark, which can be any size, location and intensity, denotes the source of the information or the identity of the requester. A watermark may be applied by the client or by the library server whenever it is wanted.

    Another technique called fingerprinting will be used to identify a user who has compromised a protected holding.

  – Encryption

    Encryption and key management techniques can be used to protect holdings. Some encryption techniques include scrambling the data itself and using a special envelope, or Cryptolope. Encryption is important for high value, or confidential, assets.

    The Digital Library resource manager will use the Cryptolope technique to transmit encrypted data. All the parts of a holding are packed together in a byte-stream (the Cryptolope) for sending to a client, where they are unpacked and protected (for example, marked) before viewing.

For export purposes, customers can install and use different cryptography techniques.

- Metering

  Metering techniques monitor the number of concurrent users of a holding, or the number of copies distributed.

  The Digital Library resource manager is able to monitor the usage of holdings in real time, and in detail, so that proper billing can be done.

- Usage fees

  The Digital Library resource manager maintains usage data on each part of a holding, and provides these usage reports to billing applications or other business processes.

The Digital Library resource manager provides the enabling software to build library solutions which can manage the unique licensing and compensation agreements required by an individual library.

Distributed Libraries

The Digital Library resource manager consists of a client portion and a server portion. The client and server portions can be configured to support a variety of configurations (see Figure 4 on page 10).

Library Clients

The client portion of the Digital Library resource manager handles requests for library functions. It supports the Digital Library resource manager programming interface. The client portion can be used on a user's workstation to support a customized solution that can access holdings from one or more digital libraries. However, a library that is supported through the Digital Library resource manager can also serve clients who are:

- WWW browsers
- Collaborative work-group clients

In all cases, the server portion of the Digital Library resource manager resides in the library server system. The library server system acts as the data store for the information being accessed by the client.

For a solution that supports a customized client, the client portion is in the user's workstation.

For the collaborative client and the WWW browser, the client portion of the Digital Library resource manager is associated with the collaboration server or HTTP server. In those cases, the user's workstation contains a collaborative client or a WWW browser.

When the client portion of the Digital Library resource manager is associated with the collaboration server or the HTTP server, it performs some unique mapping functions to provide access to the library. Therefore, the client portion acts as a digital library (DL) gateway. That is, it functions as a gateway between the collaboration or HTTP server and digital libraries.

The gateways are used to access holdings that are stored and indexed by library servers:

- When available, the HTTP DL gateway will provide forms in the Hypertext Markup Language (HTML). These forms are used in any WWW browser to compose queries. At the HTTP server, the DL gateway interprets the forms and dispatches the search requests to the library servers using the library client interface. Search parameters may include any combination of the search technologies supported by the Digital Library resource manager. See “Search and Access” on page 6.

  The results returned by the library servers are listed in a ranked order of relevance using HTML, and returned to the WWW browser. The list includes hyperlinks to the holdings.

Using the hyperlinks, users can retrieve the individual holdings. The holdings can be viewed using the web browser or an application associated with the type of the holding.

- A description of the functions of the collaboration DL gateway is not yet available.

The programming interface description in this paper tells how to build custom clients (see “Library Client Interface” on page 11). DL gateways are developed using the same interface.

Library Servers

The library server consists of a card catalog, with indexes and search support data, and one or more collections of library holdings.

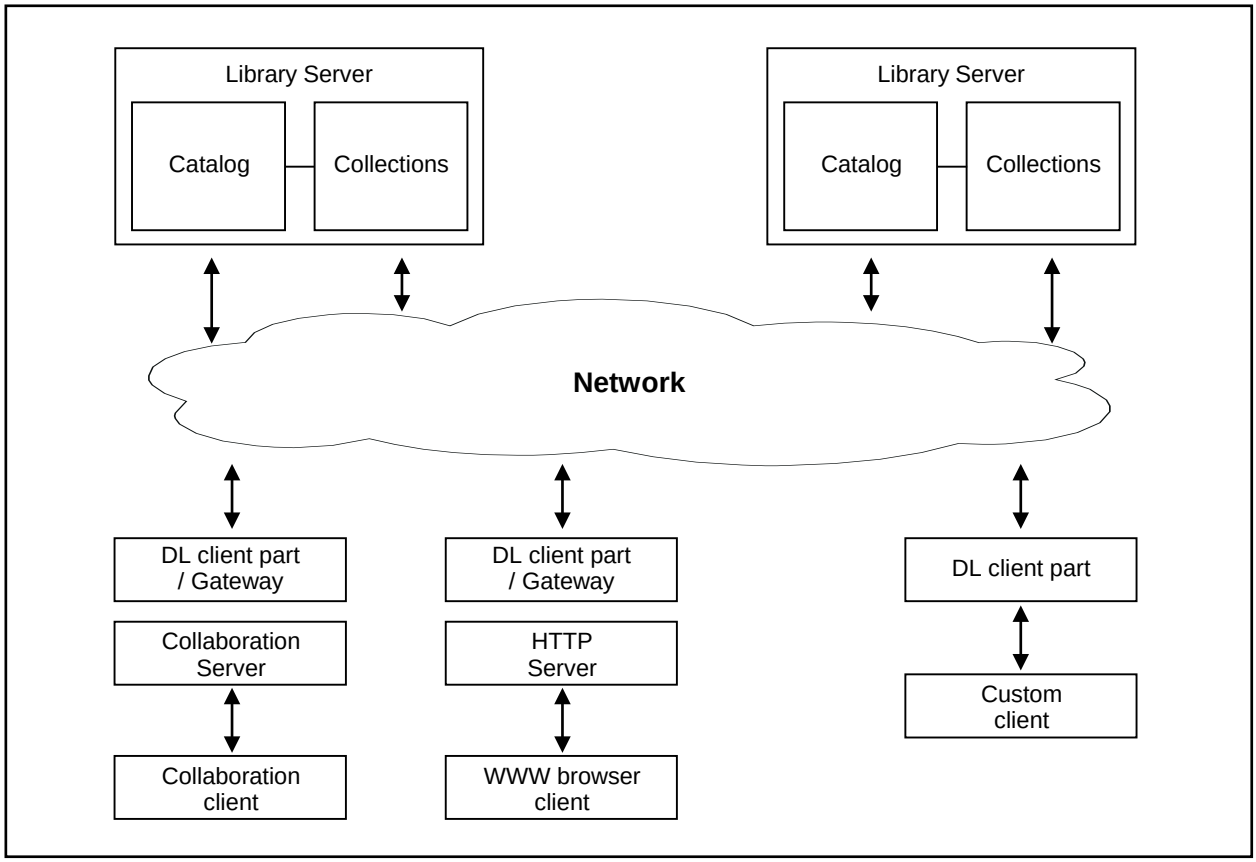


Figure 4. Distributed Libraries

The library server provides data integrity by:

- Managing library data

  The library server maintains the meta-data about the holdings in the library, as well as the holdings themselves. The meta-data are the index and search criteria (attributes) for the holdings in the collection. This includes the content-based search criteria.

  The library server uses relational database technology to manage the index and search criteria.

- Controlling access to holdings in the collection

  Each collection is associated with at least one library server. Access to the holdings in the collections are through the library server. Collections may be stored anywhere. The holdings in the collection are accessed through index information in the library server.

Any type of digitized content can be stored in a collection. The library knows nothing about the contents of the holdings it manages. However, the type of the holding must be defined to the library by a library administrator before it can be stored in the collection. Library administrators also specify *views* of the holdings and a user is authorized to see one or more views. The view determines which information about the holding a user is allowed to see.

Holdings do not have to be available online, but most scenarios assume that they are. These holdings are maintained on servers that support attachment of direct access storage devices and a range of optical storage devices. Transparent hierarchical storage management automatically moves seldom used holdings to more cost-effective secondary storage, such as optical storage.

The catalogs and the collections can reside on the same system or different systems. Collections can be added to the configuration of a library at any time. This allows a digital library to be configured to meet a wide range of storage and performance requirements. A digital library solution can scale from a single workstation, to a LAN, to an enterprise-wide solution, and to global interconnections that support distributed libraries with millions of holdings with petabytes[1] of data.

The library server is designed to work efficiently so that the actual digitized information in the collection is accessed only when requested by the user.

Library Client Interface

The programming interface supported by the library client provides the software to implement custom Digital Library resource manager clients. Library client applications, working with the Digital Library resource manager, provide broad library solutions that can be integrated with content capture and creation processes, with billing applications, and with other business processes.

Extensions to the Digital Library resource manager interface for managing collections, for rights management services such as visible watermarking, and other functions of the Digital Library resource manager are being defined.

The purpose of this section is to illustrate how a subset of the programming interface can be used for:

- Starting and ending library sessions
- Accessing library information
  - Store items in the library
  - Retrieve items from the library
  - Delete items
  - Update the attributes of an item
  - Query attributes about the items

Starting and Ending a Library Session: Figure 5 on page 12 shows a sequence of operations for starting and ending a library session.

[1] A petabyte is 10^{15}, or 1000000000000000.

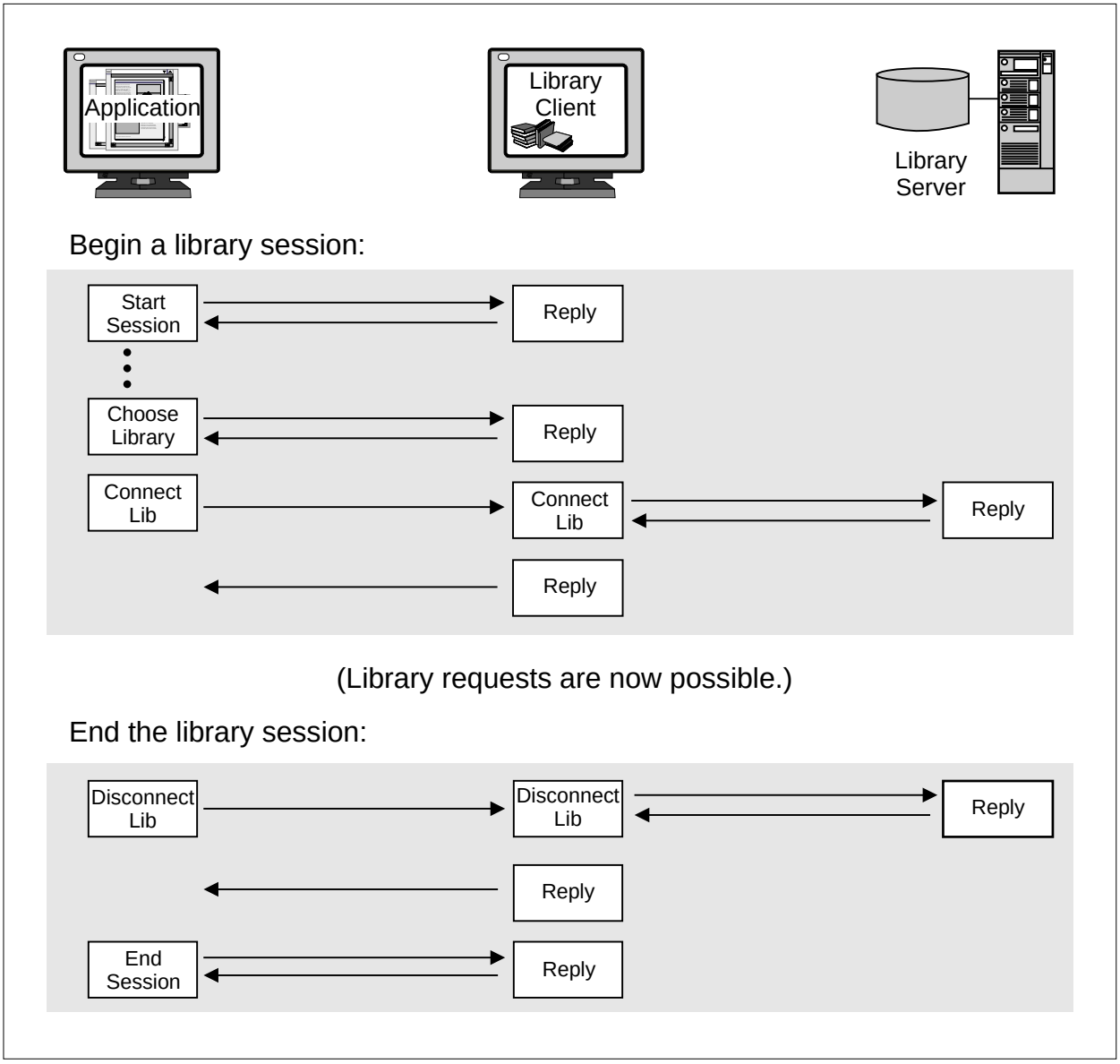


Figure 5. Starting and Ending a Library Session

Applications start the client half of a library session by choosing the **StartSession** function. From within a session, the client can choose a library (**ChooseLibrary**) and then connect to it (**ConnectLib**). The server half of the session is created when the connection request is received at the library server. When the connection is made, the library completes its half of the connection, and allows requests for other library client functions.

Client applications can disconnect from one library (**DisconnectLib**) and choose to connect with another library within the same session, or they can connect to multiple libraries within the same session. Multi-thread applications can request duplicate sessions to avoid tangling their requests and responses when dealing with multiple libraries.

When a session is completed, the application calls the **EndSession** function.

A library client application can connect to any number of library servers by requesting duplicate sessions from the Digital Library resource manager (**DupSession**).

Accessing Library Information: Items are retrieved by searching on selected attributes defined for the type. A single query can examine parametric data as well as digitized content, such as the color of an image or specific textual content.

ItemSearch is used to search for items that satisfy a set of search conditions. Search conditions consist of (*attributeID, value*) pairs with a Boolean *AND*, *OR*, or *NOT* operator.

Application programming interface (API) requests can be grouped at the client before sending them to the library server for processing. The block of requests is sent to the library server when an **EndRequest** is issued. The library server collects the responses to each request in a block before returning them to the client. When a block of responses has been returned, the client application can view the responses one at a time.

To retrieve information, the application issues a **GetResponse** request which returns a handle to the block of responses. **GetReplyInfo** retrieves the actual results for a specified call. These requests are not valid until the block of responses is returned. Figure 6 on page 14 shows a sequence of queries and responses. In the example, the information received in the search response is used to request that a holding be copied to the local workstation. This request is made using the **RetrieveItemPart** request. This is a store and forward delivery request. The library server delivers the holding to the library client directly from the collection.

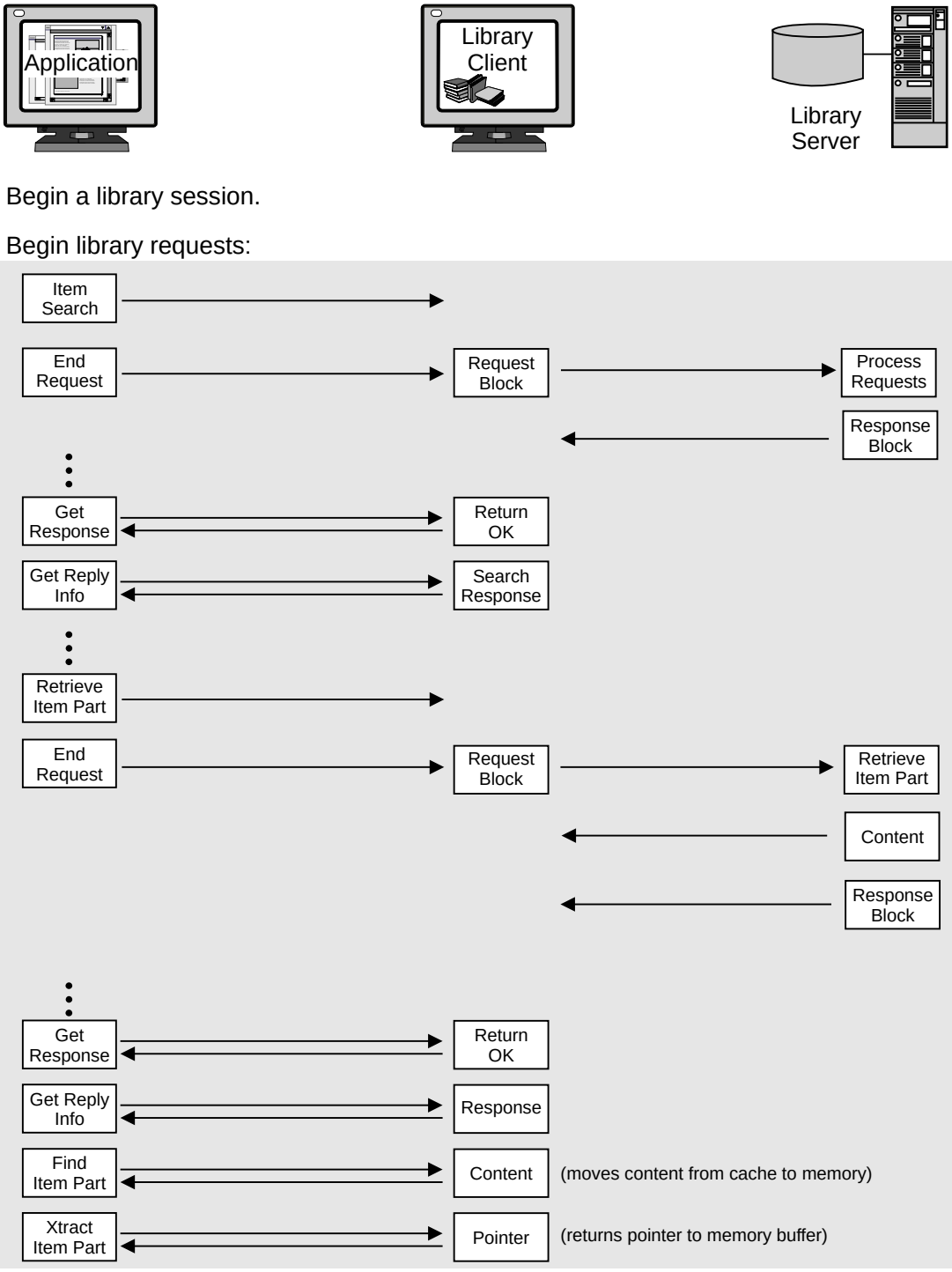


Figure 6. Accessing Library Information: Item Search

The library server also provides for both static and dynamic queries. Static queries (**StatQuery**) are faster, but dynamic queries (**DynQuery**) allow the application to change the search criteria. The static query function is used to run a predefined *SQL SELECT* statement querying the library. The dynamic query function (see Figure 7 on page 15) is used to run a dynamic *SQL SELECT* statement. The

processing flow for dynamic and static queries is the same as for **ItemSearch**. Notice that **GetQueryReplyInfo** instead of **GetReplyInfo** is used to retrieve the query data results for static and dynamic queries.

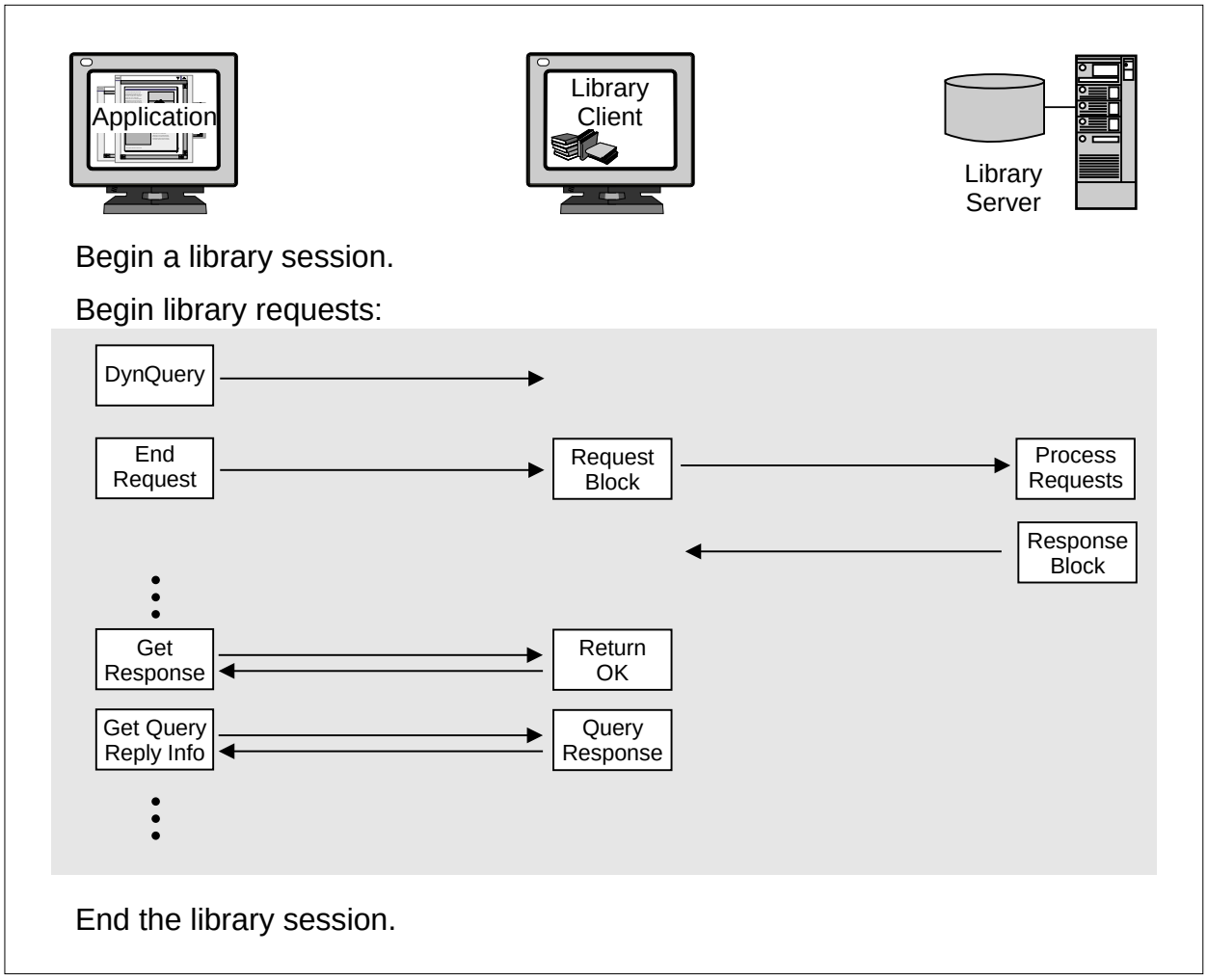


Figure 7. Accessing Library Information: Static and Dynamic Queries

Figure 8 on page 17 shows a sequence of calls, blocked into one request to the library that creates a new holding in the collection, and sets the holding's user-specified attributes in the library server. This type of request is available to add holdings to a library collection.

First, the application requests that the local cache be allocated for the digitized content using the **AllocateItemPart** function. This storage is allocated at the client.

The application then requests that the holding be stored. Because the request block is not yet completed, the data is staged temporarily in the allocated client storage.

Finally, the application sets the values of the attributes of the new holding, such as the type of the holding being stored and the identity of the owner using the **SetItemAttribute** request. This is the last functional request in this example.

Many functional requests can make up a logical unit of work. Each unit of work is ended by an **EndTransaction** request. This tells Digital Library resource manager to commit all previous requests since the last EndTransaction.

Requests can be blocked into a single request block, or they can be blocked into multiple request blocks. Several units of work can also be blocked into a single request block.

The application ends this block of requests using **EndRequest**. This causes the block of requests to be sent to the library server for processing. A new library item is created and the item attributes are set to the values specified in the **SetItemAttributes** request. The library server stores the digitized content directly in the collection.

When the processing is complete, the responses for each request are returned to the client in a response block. The application can access the response block through the **GetResponse** request, and then retrieve the reply to each individual request.

Notice that the application must free the cache when it is no longer needed (**FreeItemPart**).

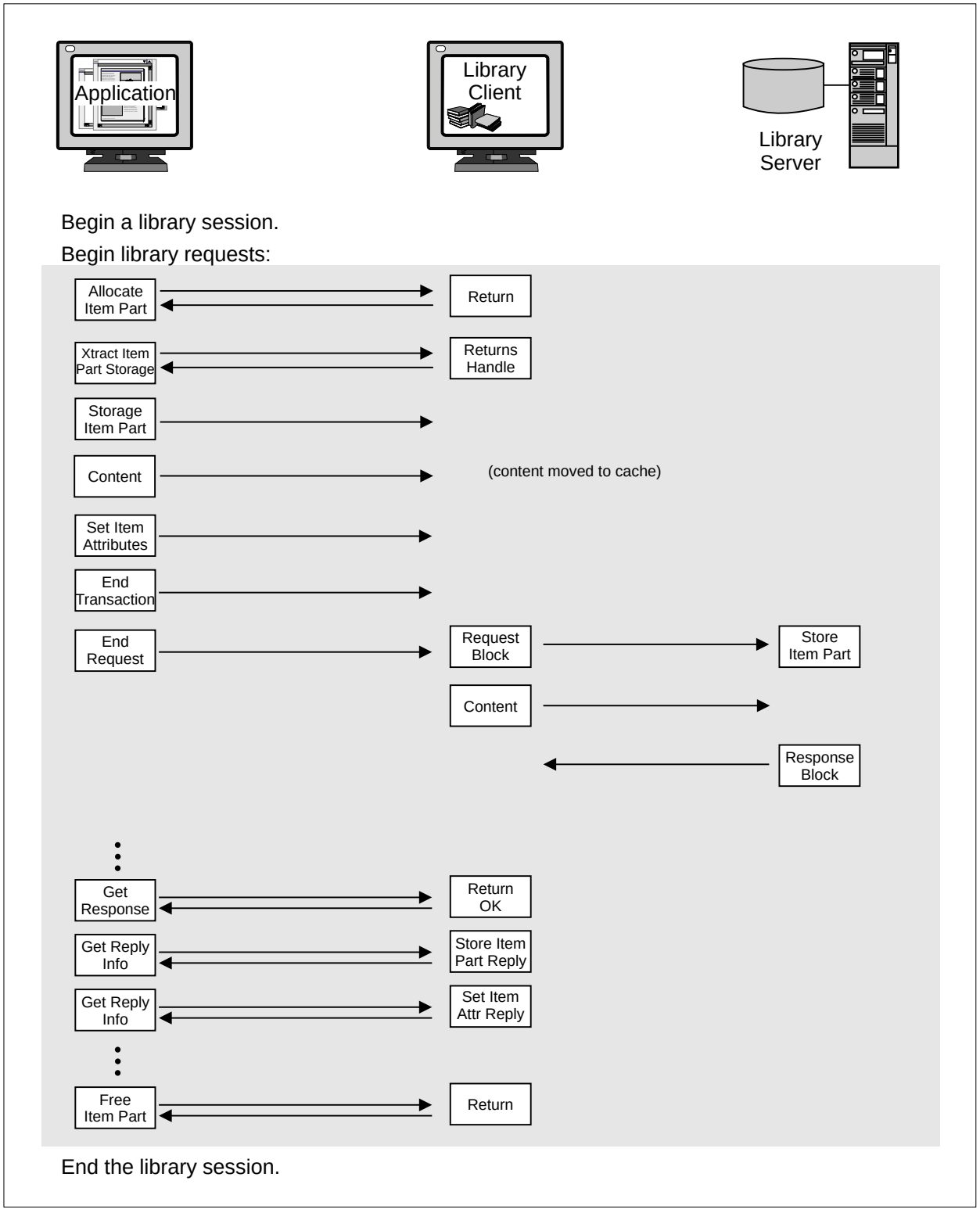


Figure 8. Accessing Library Information: Adding Holdings to a Library Collection

To prevent collisions, the library enforces concurrency procedures for updating catalog data or digitized content in the collection. Concurrency is enforced at the part level. The type of enforcement depends on whether the update is to be a short-term update or a long-term update.

Short-Term Updates: Updates that are completed in the course of a transaction are short-term updates. Concurrency is enforced by the Digital Library resource manager. Figure 8 illustrates an example of a short-term update (**StoreItemPart**, **SetItemAttributes** functions). The object is released for another update when the current transaction is committed.

Long-Term Updates: Updates that are completed across transaction boundaries are long-term updates. This happens, for example, when graphic artists need to update an image that is managed by a digital library. Because the artists need to work on the image for an extended period, perhaps days, it is not practical to keep a transaction open until the work is completed. Instead, the artists must check a holding out of the library when they intend to modify the content over an extended period. Holdings are checked out of the library using the **CheckOut** request. When the updates are completed the holding is checked back into the library using the **CheckIn** request.

Concurrency is enforced because no other user can check out the same holding until it is checked in by the original requester or released by the requester or the library administrator.

Standards

Many of the standards that apply to, and are used by, the Digital Library resource manager are provided by other components of the Open Blueprint. Other standards being considered for the Digital Library resource manager are described in this section.

Document Management Alliance

The Document Management Alliance (DMA) is a standards-setting body working on programming interfaces that relate to the Digital Library resource manager. The DMA was formed from the merger of Document Enabled Networking and the Shamrock Document Management Coalition. It is currently organized as a task force under the Association for Information and Image Management (AIIM).

A document may be anything from a computer file to a video clip. The objective of the DMA specification is to define:

1. A common interface for integrating access and search methods
2. A vendor independent API for accessing and searching across diverse asset management services
3. An object-based data model to standardize access to library services that includes version control, security, and query services

When complete, the results of this work will be evaluated for its applicability to the Digital Library resource manager.

Shared Library Access

Shared library access will be made possible by evolving standards for bibliographic records. In the 1960s, libraries began to develop standards to denote the information that must be cataloged for a new holding (the machine readable catalog (MARC) record format), and how that record could be transferred among catalogs to eliminate duplicate cataloging (the Z39.2 format). Standard rules on how to enter the information into the MARC record are known as the Anglo-American Cataloging Rules (AACR).

The term *bibliographic* should be construed literally, because current conventional library catalogs describe little more than books. However, the standards can be extended to support multimedia holdings.

Existing legacy bibliographic catalogs can reference the Digital Library resource manager. To do this, a pointer to the Digital Library resource manager holding can be added in the Electronic Location field in the existing MARC record.

The work to standardize the organization and content of information was soon followed by standards governing online access and retrieval, called the *Information Retrieval Service and Protocol Standard (ANSI Z39.50)*. This standard governs the exchange of messages between a requesting system and a responding system that enable the requester to search a library and receive the results of the search. The standard applies to the communication between the requesting system and the responding system; it does not address the interaction with the user. Z39.50 will be supported where standardized catalog queries are needed.

Relationship to Other Resource Managers

Collaboration Resource Manager

The Digital Library resource manager intends to make the holdings in a library server accessible to collaboration work group clients through a gateway in the collaboration server.

Relational Database Resource Manager: The library server requires a high level of availability, integrity, performance, and security to provide reliable service to the library owners and their users. The storage and management model of the Digital Library resource manager shown in Figure 1 on page 1 relies on the Relational Database resource manager to manage relational tables holding the index and search support attributes of the library holdings.

Storage Management Resource Manager: To optimize library resources, Digital Library resource manager transparently moves seldom accessed holdings to more cost-effective secondary storage. This is accomplished through the use of the space management interface of the Storage Management resource manager. This component automatically migrates eligible files to secondary storage in order to maintain specific levels of free space in the collection. It automatically recalls these migrated holdings when they are accessed.

Storage management policies are established by the library administrator. These policies indicate how much free space to maintain, which holdings are candidates for migration, whether or not to maintain backup copies, and the storage hierarchy in which holdings should be placed.

The library administrator can tune the operation of the space management facility to suit the needs of the library. This is transparent to the library client, and this capability is not reflected in the Digital Library resource manager interface.

HTTP Resource Manager: The Digital Library resource manager intends to make the holdings in a library server accessible to WWW browsers through a gateway in the HTTP resource manager.

Identification and Authentication Resource Manager: From time to time, the identity of a user, or a key identifying a set of users, is required. The Digital Library resource manager intends that such information will be reliably extracted from the Identification and Authentication resource manager when needed.

In the context of the Open Blueprint, a user must be authenticated and identified by the Identification and Authentication resource manager prior to making any request of the library.

The generic security services (GSS-API) interfaces with the Identification and Authentication resource manager.

Access Control Resource Manager: Access to holdings in a digital library is controlled by the digital library itself. Access and privileges are granted to a library user by the library administrator, and will differ among libraries. It is possible for a user's access and privileges to change as the user connects from one library to another.

Communications Support

The Digital Library resource manager isolates all communications functions from other Digital Library resource manager functions. The communications component supports SNA APPN and TCP/IP networks. The use of Open Blueprint Messaging and Queueing resource manager is being evaluated.

Directory Resource Manager

Library servers, collections, and holdings in the library all have names. The library server name, and the name of any server that hosts a collection, are unique by virtue of inclusion in the Open Blueprint Directory name space. The library server itself understands and supports the names of the library holdings in the collections. These holdings names are not registered in the Open Blueprint Directory.

Multimedia Resource Manager

Digital Library resource manager does not directly interface with the Multimedia resource manager, but library client applications are expected to do so in order to stream audio and video information to the client.

Transaction Support

The library server part of the Digital Library resource manager will use the Open Blueprint Transaction Monitor resource manager and Transaction Manager resource manager wherever they are available.

Digital Library Resource Manager and OO Technology

The Digital Library resource manager is evaluating a class library to support search and access functions to digital library clients based on the Object Query Services specification of the Object Management Group (OMG). The specification is in the *OMG TC Document 95.1.1*.

System Management

Digital Library resource manager supports local system management interfaces for error reporting at the client, the library server, and the collection on behalf of the resources it manages.

View Resource Manager

The Digital Library resource manager does not directly interface with the View resource manager, but library client applications are expected to do so in order to view printable documents retrieved from a library server.

Library Administration

There are three core elements in managing a digital library:

1. Managing access to the library

   For access control, library users can be granted individual privileges, or group privileges by the library administrator. Users must be identified in order to access a library. Users will be identified by the certificate returned by the Identification and Authentication resource manager. Separate logon identities will not be required. Billing and licensing information about each user is also specified by the library administrator. This information is used to monitor usage, to protect the holdings, and to determine the usage statistics which must be kept.

2. Administering holdings and collections

   The library administrator:

   - Defines the types of holdings that are managed by the library, and the views of each type of holding. It is the *view* that is referenced rather than the type definition in subsequent client operations.
   - Specifies the billing and licensing information about each holding. Holdings cannot be added to the library until the type of the holding has been defined to the library server.
   - Specifies the index and search criteria for library holdings.

3. Administering library storage

   The administrator defines system-managed library storage in the collections, partitions the various holdings among the collections, and assigns collections to library servers. Collections can be added, changed or removed from a library at any time.

   The administrator provides a name for the library and specifies the logon policies, library server rules, and access control defaults.

   Logon policies include the maximum number of concurrent users, timeout criteria, and so on.

   Library server rules include code page use, update criteria, memory usage, maximum checkout times, and more.

   The Storage Management resource manager is used to provide transparent migration of seldom used holdings to secondary storage. The Digital Library resource manager library administrator has tools to monitor this activity. These tools provide statistics needed to help establish migration policies.

Notices

References in this publication to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Subject to IBM's valid intellectual property or other legally protectable rights, any functionally equivalent product, program, or service may be used instead of the IBM product, program, or service. Evaluation and verification of operation in conjunction with other products, except those expressly designated by IBM, is the user's responsibility.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
500 Columbus Avenue
Thornwood, NY 10594
USA

Trademarks

The following terms are trademarks of the IBM Corporation in the United States or other countries or both:

APPN
Cryptolope
IBM
IBMLink
Open Blueprint

The following terms are trademarks of other companies:

Prodigy	Prodigy Services Company

Communicating Your Comments to IBM

If you especially like or dislike anything about this paper, please use one of the methods listed below to send your comments to IBM. Whichever method you choose, make sure you send your name, address, and telephone number if you would like a reply. Feel free to comment on specific error or omissions, accuracy, organization, subject matter, or completeness of this paper.

When you send comments to IBM, you grant IBM a nonexclusive right to use or distribute your comments in any way it believes appropriate without incurring any obligation to you.

- If you prefer to send comments by FAX, use this number:

  United States and Canada: 1-800-227-5088.

- If you prefer to send comments electronically, use one of these ID's:
  - Internet: **USIB2HPD@VNET.IBM.COM**
  - IBM Mail Exchange: **USIB2HPD at IBMMAIL**
  - IBMLink: **CIBMORCF at RALVM13**

Make sure to include the following in your note:

- Title of this paper
- Page number or topic to which your comment applies

Printed in the United States of America
on recycled paper containing 10%
recovered post-consumer fiber.

GC23-3909-01